

## RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10|S06, 651

Source: PCT

Date Processed by STIC: 11-24-04

# ***ENTERED***



PCT

## RAW SEQUENCE LISTING

DATE: 11/24/2004

PATENT APPLICATION: US/10/506,651

TIME: 12:24:16

Input Set : A:\Nih-458.app

Output Set: N:\CRF4\11242004\J506651.raw

```

3 <110> APPLICANT: Dimitrov, Dimiter Stanchev
4     Chow, Yen-Hung
5     Phogat, Sanjay Kumar
6     Broder, Christopher Charles
7     The Government of the United States of America
8     as represented by The Secretary of the
9     Department of Health and Human Services
10 <120> TITLE OF INVENTION: HIV-1 Envelope Glycoproteins Stabilized by Flexible
11     Linkers as Potent Entry Inhibitors and Immunogens
12 <130> FILE REFERENCE: 015280-458000US
13 <140> CURRENT APPLICATION NUMBER: US 10/506,651
14 <141> CURRENT FILING DATE: 2004-09-02
15 <150> PRIOR APPLICATION NUMBER: WO PCT/US02/07144
16 <151> PRIOR FILING DATE: 2002-03-05
17 <160> NUMBER OF SEQ ID NOS: 33
18 <170> SOFTWARE: PatentIn Ver. 2.1
19 <210> SEQ ID NO: 1
20 <211> LENGTH: 2562
21 <212> TYPE: DNA
22 <213> ORGANISM: Human immunodeficiency virus type 1
23 <220> FEATURE:
24 <223> OTHER INFORMATION: full length human immunodeficiency virus-1 (HIV-1)
25     envelope glycoprotein (Env, gp160) from primary
26     R5X4 isolate 89.6
27 <400> SEQUENCE: 1
28 atgagagtga aggagatcag gaagaattgg cagcacttga gagggggcat cttgctcctt 60
29 gggatgttga tgatctgtag tgctgcaaaa gaaaaaacgt gggtcacaat ctattatggg 120
30 gtacctgtgt ggagagaagc aaccaccact ctattttgtg catcagatgc taaagcctat 180
31 gatacagagg tacataatgt ttgggccaca catgcctgtg taccacaga cccaaccca 240
32 caagaagtag tattgggaaa tgtgacagaa aattttaaca tgtggaaaaa taacatggta 300
33 gatcagatgc atgaggatat aatcagttta tgggatgaaa gcctaaagcc atgtgtaaaa 360
34 ttaacccac tctgtgttac tttaaattgc actaatttga atatcactaa gaatactact 420
35 aatcccacta gtagcagctg gggaatgatg gagaaaggag aaataaaaaa ttgctctttc 480
36 tatatcacca caagcataag aaataaggta aagaaagaat atgcactttt taatagactt 540
37 gatgtagtac caatagaaaa tactaataat actaagtata ggtaataaag ttgtaacacc 600
38 tcagtcatta cacaggcctg tccaaaggta tcctttcagc caattcccat acattattgt 660
39 gtcccggctg ggtttgcgat gctaaagtgt aacaataaga cattcaatgg atcaggacca 720
40 tgcacaaatg tcagcacagt acaatgtaca catggaatta ggccagtggg gtcaactcaa 780
41 ctgctgttaa atggcagctc agcagaagaa gacatagtaa ttagatctga aaatttcaca 840
42 gacaatgcta aaaccataat agtacagcta aatgaatctg tagtaattaa ttgtacaaga 900
43 cccaacaaca atacaagaag aaggttatct ataggaccag ggagagcatt ttatgcaaga 960
44 agaaacataa taggagatat aagacaagca cattgtaaca ttagtagagc aaaatggaat 1020
45 aacactttac aacagatagt tataaaatta agagaaaaat ttaggaataa aacaatagcc 1080

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/506,651

DATE: 11/24/2004

TIME: 12:24:16

Input Set : A:\Nih-458.app

Output Set: N:\CRF4\11242004\J506651.raw

```

55 tttaaatcaat cctcaggagg ggacccagaa attgtaatgc acagtttttaa ttgtggaggg 1140
56 gaatttttct actgtaatac agcacaactg tttaatagta cttggaatgt tactggaggg 1200
57 acaaatggca ctgaaggaaa tgacataatc aactccaat gcagaataaa acaaattata 1260
58 aatatgtggc agaaagtagg aaaagcaatg tatgcccctc ccatcacagg acaaattaga 1320
59 tgttcatcaa atattacagg gctgctacta acaagagatg gaggtaatag tactgagact 1380
60 gagactgaga tcttcagacc tggaggagga gatatgaggg acaattggag aagtgaatta 1440
61 tataaatata aagtagtaag aattgaacca ataggagtag caccaccag ggcaaagaga 1500
62 agaacagtgc aaagagaaaa aagagcagtg ggaaataggag ctgtgttcct tgggttcttg 1560
63 ggagcagcag gaagcactat gggcgagcg tcagtgcgc tgacgggtaca ggccaggcta 1620
64 ttattgtctg gtatagtgc gcagcagaac aatctgctga gggctattga ggcgcaacag 1680
65 catatgttgc aactcacagt ctggggcatc aagcagctcc aggcaagagt cctggctctg 1740
66 gaaagatacc taagggatca acagctcatg ggaatttggg gttgctctgg aaaactcatt 1800
67 tgcaccactt ctgtgccttg gaatgttagt tggagtaata aatctgtgga tgatatttgg 1860
68 aataacatga cctggatgga gtgggaaaga gaaattgaca attacacaga ctatatatat 1920
69 gacttacttg aaaaatcgca aaccacaaca gaaaagaatg aaaaagaatt attggaattg 1980
70 gataaatggg caagtttgtg gaattggttt gacataacaa actggctgtg gtatataaga 2040
71 ttattcataa tgatagtagg aggcttgata ggtttaagaa tagtttttgc tgtactttct 2100
72 atagtaaata gagttaggca gggatattca ccattatcgt ttcagaccct cctcccagcc 2160
73 tgcaggggac ccgacaggcc cgaaggaaca gaagaagaag gtggagagag agacagagac 2220
74 agatccggtc cattagtga cggattcttg gcacttttct gggtcgattt gaggaacctg 2280
75 tgcctcttcc tctaccacct cttgagaaac ttactcttga ttgtaacgag gattgtggaa 2340
76 cttctgggac gcaggggggtg ggaagccctc aatatattgt ggaatctcct gcaatattgg 2400
77 agccaggaac taaagaatag tgctgttagc ttgctcaacg ccacagccat agcagtagct 2460
78 gaggggacag atagggttat aaaaatagta caaagagctt gtagagctat tcgcaacata 2520
79 cctacaagaa tcagacaggg cttggaaagg gctttgctat aa 2562
82 <210> SEQ ID NO: 2
83 <211> LENGTH: 853
84 <212> TYPE: PRT
85 <213> ORGANISM: Human immunodeficiency virus type 1
87 <220> FEATURE:
88 <223> OTHER INFORMATION: full length human immunodeficiency virus-1 (HIV-1)
89 envelope glycoprotein (Env, gp160) from primary
90 R5X4 isolate 89.6
92 <400> SEQUENCE: 2
93 Met Arg Val Lys Glu Ile Arg Lys Asn Trp Gln His Leu Arg Gly Gly
94 1 5 10 15
96 Ile Leu Leu Leu Gly Met Leu Met Ile Cys Ser Ala Ala Lys Glu Lys
97 20 25 30
99 Thr Trp Val Thr Ile Tyr Tyr Gly Val Pro Val Trp Arg Glu Ala Thr
100 35 40 45
102 Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
103 50 55 60
105 His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
106 65 70 75 80
108 Gln Glu Val Val Leu Gly Asn Val Thr Glu Asn Phe Asn Met Trp Lys
109 85 90 95
111 Asn Asn Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
112 100 105 110
114 Glu Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/506,651

DATE: 11/24/2004

TIME: 12:24:16

Input Set : A:\Nih-458.app

Output Set: N:\CRF4\11242004\J506651.raw

115	115	120	125
117 Asn Cys Thr Asn Leu Asn Ile Thr Lys Asn Thr Thr Asn Pro Thr Ser			
118 130	135	140	
120 Ser Ser Trp Gly Met Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe			
121 145	150	155	160
123 Tyr Ile Thr Thr Ser Ile Arg Asn Lys Val Lys Lys Glu Tyr Ala Leu			
124 165	170	175	
126 Phe Asn Arg Leu Asp Val Val Pro Ile Glu Asn Thr Asn Asn Thr Lys			
127 180	185	190	
129 Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro			
130 195	200	205	
132 Lys Val Ser Phe Gln Pro Ile Pro Ile His Tyr Cys Val Pro Ala Gly			
133 210	215	220	
135 Phe Ala Met Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Ser Gly Pro			
136 225	230	235	240
138 Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val			
139 245	250	255	
141 Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Asp Ile			
142 260	265	270	
144 Val Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val			
145 275	280	285	
147 Gln Leu Asn Glu Ser Val Val Ile Asn Cys Thr Arg Pro Asn Asn Asn			
148 290	295	300	
150 Thr Arg Arg Arg Leu Ser Ile Gly Pro Gly Arg Ala Phe Tyr Ala Arg			
151 305	310	315	320
153 Arg Asn Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Arg			
154 325	330	335	
156 Ala Lys Trp Asn Asn Thr Leu Gln Gln Ile Val Ile Lys Leu Arg Glu			
157 340	345	350	
159 Lys Phe Arg Asn Lys Thr Ile Ala Phe Asn Gln Ser Ser Gly Gly Asp			
160 355	360	365	
162 Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr			
163 370	375	380	
165 Cys Asn Thr Ala Gln Leu Phe Asn Ser Thr Trp Asn Val Thr Gly Gly			
166 385	390	395	400
168 Thr Asn Gly Thr Glu Gly Asn Asp Ile Ile Thr Leu Gln Cys Arg Ile			
169 405	410	415	
171 Lys Gln Ile Ile Asn Met Trp Gln Lys Val Gly Lys Ala Met Tyr Ala			
172 420	425	430	
174 Pro Pro Ile Thr Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu			
175 435	440	445	
177 Leu Leu Thr Arg Asp Gly Gly Asn Ser Thr Glu Thr Glu Thr Glu Ile			
178 450	455	460	
180 Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu			
181 465	470	475	480
183 Tyr Lys Tyr Lys Val Val Arg Ile Glu Pro Ile Gly Val Ala Pro Thr			
184 485	490	495	
186 Arg Ala Lys Arg Arg Thr Val Gln Arg Glu Lys Arg Ala Val Gly Ile			
187 500	505	510	

## RAW SEQUENCE LISTING

DATE: 11/24/2004

PATENT APPLICATION: US/10/506,651

TIME: 12:24:16

Input Set : A:\Nih-458.app

Output Set: N:\CRF4\11242004\J506651.raw

```

189 Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly
190      515      520      525
192 Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg Leu Leu Leu Ser Gly
193      530      535      540
195 Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln
196 545      550      555      560
198 His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg
199      565      570      575
201 Val Leu Ala Leu Glu Arg Tyr Leu Arg Asp Gln Gln Leu Met Gly Ile
202      580      585      590
204 Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ser Val Pro Trp Asn
205      595      600      605
207 Val Ser Trp Ser Asn Lys Ser Val Asp Asp Ile Trp Asn Asn Met Thr
208      610      615      620
210 Trp Met Glu Trp Glu Arg Glu Ile Asp Asn Tyr Thr Asp Tyr Ile Tyr
211 625      630      635      640
213 Asp Leu Leu Glu Lys Ser Gln Thr Gln Gln Glu Lys Asn Glu Lys Glu
214      645      650      655
216 Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp Ile
217      660      665      670
219 Thr Asn Trp Leu Trp Tyr Ile Arg Leu Phe Ile Met Ile Val Gly Gly
220      675      680      685
222 Leu Ile Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn Arg
223      690      695      700
225 Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Leu Leu Pro Ala
226 705      710      715      720
228 Ser Arg Gly Pro Asp Arg Pro Glu Gly Thr Glu Glu Glu Gly Gly Glu
229      725      730      735
231 Arg Asp Arg Asp Arg Ser Gly Pro Leu Val Asn Gly Phe Leu Ala Leu
232      740      745      750
234 Phe Trp Val Asp Leu Arg Asn Leu Cys Leu Phe Leu Tyr His Leu Leu
235      755      760      765
237 Arg Asn Leu Leu Leu Ile Val Thr Arg Ile Val Glu Leu Leu Gly Arg
238      770      775      780
240 Arg Gly Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr Trp
241 785      790      795      800
243 Ser Gln Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn Ala Thr Ala
244      805      810      815
246 Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Lys Ile Val Gln Arg
247      820      825      830
249 Ala Cys Arg Ala Ile Arg Asn Ile Pro Thr Arg Ile Arg Gln Gly Leu
250      835      840      845
252 Glu Arg Ala Leu Leu
253      850
256 <210> SEQ ID NO: 3
257 <211> LENGTH: 2051
258 <212> TYPE: DNA
259 <213> ORGANISM: Human immunodeficiency virus type 1
261 <220> FEATURE:

```

## RAW SEQUENCE LISTING

DATE: 11/24/2004

PATENT APPLICATION: US/10/506,651

TIME: 12:24:16

Input Set : A:\Nih-458.app

Output Set: N:\CRF4\11242004\J506651.raw

```

262 <223> OTHER INFORMATION: human immunodeficiency virus-1 (HIV-1) envelope
263     glycoprotein gp140 truncated version of gp 160
264     from primary R5X4 isolate 89.6
266 <400> SEQUENCE: 3
267 atgagagtga aggagatcag gaagaattgg cagcacttga gaggggggcat cttgctcctt 60
268 gggatgttga tgatctgtag tgctgcaaaa gaaaaaacgt gggtcacaat ctattatggg 120
269 gtacctgtgt ggagagaagc aaccaccact ctattttgtg catcagatgc taaagcctat 180
270 gatacagagg tacataatgt ttggggccaca catgcctgtg taccacacaga cccaaccca 240
271 caagaagtag tattgggaaa tgtgacagaa aattttaaca tgtggaaaaa taacatggta 300
272 gatcagatgc atgaggatat aatcagttta tgggatgaaa gcctaaagcc atgtgtaaaa 360
273 ttaaccccac tctgtgttac tttaaattgc actaatttga atatcactaa gaatactact 420
274 aatcccacta gtagcagctg gggaatgatg gagaaaggag aaataaaaaa ttgctctttc 480
275 tatatcacca caagcataag aaataaggta aagaaagaat atgcactttt taatagactt 540
276 gatgtagtac caatagaaaa tactaataat actaagtata ggtaataaag ttgtaacacc 600
277 tcagtcatta cacaggcctg tccaaaggta tcctttcagc caattcccat acattattgt 660
278 gtcccggctg ggtttgcgat gctaaagtgt aacaataaga cattcaatgg atcaggacca 720
279 tgcacaaatg tcagcacagt acaatgtaca catggaatta ggccagtggg gtcaactcaa 780
280 ctgctgttaa atggcagctc agcagaagaa gacatagtaa ttagatctga aaatttcaca 840
281 gacaatgcta aaaccataat agtacagcta aatgaatctg tagtaattaa ttgtacaaga 900
282 cccaacaaca atacaagaag aaggttatct ataggaccag ggagagcatt ttatgcaaga 960
283 agaaacataa taggagatat aagacaagca cattgtaaca ttagtagagc aaaatggaat 1020
284 aacactttac aacagatagt tataaaatta agagaaaaat ttaggaataa aacaatagcc 1080
285 tttaatcaat cctcaggagg ggacccagaa attgtaatgc acagttttaa ttgtggaggg 1140
286 gaatttttct actgtaatac agcacaactg tttaatagta cttggaatgt tactggaggg 1200
287 acaaatggca ctgaaggaaa tgacataatc acactccaat gcagaataaa acaaatata 1260
288 aatatgtggc agaaagtagg aaaagcaatg tatgcccctc ccatcacagg acaaatata 1320
289 tgttcatcaa atattacagg gctgctacta acaagagatg gaggtaatag tactgagact 1380
290 gagactgaga tcttcagacc tggaggagga gatatgaggg acaattggag aagtgaatta 1440
291 tataaatata aagtagtaag aattgaacca ataggagtag caccaccag ggcaaagaga 1500
292 agaacagtgc aaagagaaaa aagagcagtg ggaataggag ctgtgttcct tgggttcttg 1560
293 ggagcagcag gaagcactat gggcgcagcg tcagtgcacg tgacgggtaca ggccaggcta 1620
294 ttattgtctg gtatagtgcg gcagcagaac aatctgctga gggctattga ggcgcaacag 1680
295 catatgttgc aactcacagt ctggggcctc aagcagctcc aggcaagagt cctggctctg 1740
296 gaaagatacc taagggatca acagctcatg ggaatttggg gttgctctgg aaaactcatt 1800
297 tgcaccactt ctgtgccttg gaatgttagt tggagtaata aatctgtgga tgatatttgg 1860
298 aataacatga cctggatgga gtgggaaaga gaaattgaca attacacaga ctatatatat 1920
299 gacttacttg aaaaatcgca aaccaacaa gaaaagaatg aaaaagaatt attggaattg 1980
300 gataaatggg caagtttgtg gaattggttt gacataacaa actggctgtg gtatataaga 2040
301 ttattcataa t 2051
304 <210> SEQ ID NO: 4
305 <211> LENGTH: 667
306 <212> TYPE: PRT
307 <213> ORGANISM: Human immunodeficiency virus type 1
309 <220> FEATURE:
310 <223> OTHER INFORMATION: human immunodeficiency virus-1 (HIV-1) envelope
311     glycoprotein gp140 truncated version of gp 160
312     from primary R5X4 isolate 89.6
314 <400> SEQUENCE: 4
315 Met Arg Val Lys Glu Ile Arg Lys Asn Trp Gln His Leu Arg Gly Gly

```

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/506,651

DATE: 11/24/2004

TIME: 12:24:17

Input Set : A:\Nih-458.app

Output Set: N:\CRF4\11242004\J506651.raw